



Quantitative Profiling of the Activity of Protein Lysine Methyltransferase SMYD2 Using SILAC-Based Proteomics^{*}

Jonathan B. Olsen^{‡**}, Xing-Jun Cao^{§**}, Bomie Han[‡], Lisa Hong Chen[‡], Alexander Horvath[¶], Timothy I. Richardson[‡], Robert M. Campbell[‡], Benjamin A. Garcia^{§||}, and Hannah Nguyen^{‡||}

The significance of non-histone lysine methylation in cell biology and human disease is an emerging area of research exploration. The development of small molecule inhibitors that selectively and potently target enzymes that catalyze the addition of methyl-groups to lysine residues, such as the protein lysine mono-methyltransferase SMYD2, is an active area of drug discovery. Critical to the accurate assessment of biological function is the ability to identify target enzyme substrates and to define enzyme substrate specificity within the context of the cell. Here, using stable isotopic labeling with amino acids in cell culture (SILAC) coupled with immunoaffinity enrichment of mono-methyl-lysine (Kme1) peptides and mass spectrometry, we report a comprehensive, large-scale proteomic study of lysine mono-methylation, comprising a total of 1032 Kme1 sites in esophageal squamous cell carcinoma (ESCC) cells and 1861 Kme1 sites in ESCC cells overexpressing SMYD2. Among these Kme1 sites is a subset of 35 found to be potently down-regulated by both shRNA-mediated knockdown of SMYD2 and LLY-507, a selective small molecule inhibitor of SMYD2. In addition, we report specific protein sequence motifs enriched in Kme1 sites that are directly regulated by endogenous SMYD2 activity, revealing that SMYD2 substrate specificity is more diverse than expected. We further show direct activity of SMYD2 toward BTF3-K2, PDAP1-K126 as well as numerous sites within the repetitive units of two unique and exceptionally large proteins, AHNAK and AHNAK2. Collectively, our findings provide quantitative insights into the cellular activity and substrate recognition of SMYD2 as well as the global landscape and regulation of protein mono-methylation. *Molecular &*

Cellular Proteomics 15: 10.1074/mcp.M115.053280, 892–905, 2016.

Protein lysine methyltransferases (PKMTs)¹ catalyze the sequence-specific transfer of one, two, or three methyl groups to the side chains of lysine residues (1–4). In addition to the extensively studied lysine methylation on histones, PKMTs can modify non-histone proteins (3, 5–8). An increasing number of non-histone proteins have been reported as PKMT substrates, and, as a result, potential roles for the dysregulation of non-histone lysine methylation in cancer development and progression have been proposed (9). The majority of PKMT substrates have been identified based on biochemical methylation assays using recombinant enzyme and substrate, followed by cell-based assays typically requiring overexpression of enzyme and/or substrate to maximize signal detection (10–13). However, it is difficult to discern whether these substrates are *bona fide* physiological substrates of endogenous PKMT activity. With the development of PKMT-targeted drugs emerging as a key area of drug discovery (14–18), unbiased and quantitative methods enabling the comprehensive identification of histone and non-histone substrates in cells are critical to clarifying the cell-relevant substrates of these enzymes and to more accurately understand the functions of non-histone methylation.

Progressing from targeted biochemical assays for substrate identification toward global and unbiased assays that monitor PKMT activity in cells requires advanced proteomic approaches that can identify methylation sites in sufficient depth and coverage and accurately quantify changes in the abundance of these sites in response to perturbation. Mass spectrometry-based proteomics is a well-established tool for the

From the [‡]Lilly USA, Lilly Research Laboratories, Indianapolis, Indiana 46285; [§]Epigenetics Program, Department of Biochemistry and Biophysics, Perelman School of Medicine, University of Pennsylvania Philadelphia, Pennsylvania 19104; [¶]Advanced Testing Laboratory, Cincinnati, Ohio, 45242

Received July 14, 2014, and in revised form, January 4, 2016

Published, MCP Papers in Press, January 10, 2016, DOI 10.1074/mcp.M115.053280

Author contributions: J.B.O., X.C., B.H., T.I.R., R.M.C., B.A.G., and H.N. designed research; J.B.O., X.C., B.H., L.H.C., and A.H. performed research; J.B.O., X.C., and B.H. analyzed data; J.B.O., X.C., B.A.G., and H.N. wrote the paper.

¹ The abbreviations used are: PKMT, Protein lysine methyltransferase; CRU, Central repeat unit; DMSO, Dimethyl sulfoxide; ESCC, Esophageal squamous cell carcinoma; HPLC, High-performance liquid chromatography; Kme1, Mono-methyl lysine; LC-MS/MS, Liquid chromatography and tandem mass spectrometry; PSM, Peptide-spectrum match; PTM, Post-translational modification; SCX, Strong cation exchange; SILAC, Stable isotopic labelling with amino acids in cell culture.

global identification of post-translational modification (PTM) sites and can be effectively coupled with immunoaffinity enrichment of peptides using “pan-specific” antibodies that bind to modified peptides independent of surrounding sequence context (19, 20). Enrichment of modified peptides using pan-specific antibodies has allowed for the identification of thousands of acetylated and ubiquitinated lysine sites in cells (21–23), and we and others have shown that pan-specific methyl-lysine or methyl-arginine antibodies can effectively and robustly enrich methylation modifications as well (7, 8). Immunoaffinity enrichment is particularly critical in the case of methyl-lysine peptides, as the methyl moiety itself imparts sufficiently little physiochemical change to the side chain, rendering the enrichment of these peptides refractory to conventional separation methods. Other approaches that exploit the naturally-occurring binding specificity of methyl-lysine recognition protein domains, such as chromodomains and MBT domains, have also shown some promise as enrichment reagents (24, 25). Nevertheless, the enrichment of non-histone methyl-lysine peptides has not yet reached the success that other types of PTMs (*i.e.* phosphorylation, ubiquitination, GlcNAc, etc.) have met in recent years. Additionally, there are currently no comprehensive large-scale proteomic studies that have quantitatively measured the relative changes in methyl-lysine site abundance or that link the regulation of specific non-histone methyl-lysine sites to the activity of a PKMT or demethylase.

SMYD2 is a cytoplasmic PKMT that is overexpressed or amplified in various types of cancers (17, 26, 27), suggesting potential roles for the mono-methyltransferase in oncogenesis. Biochemical substrates of SMYD2 have been described, most notably the tumor suppressors p53 (10) and retinoblastoma (RB1) (13), and p53 has formed the basis of biochemical and cell-based assays that have guided the development of selective cell-permeable inhibitors of SMYD2 (28, 29), including LLY-507 (17). Biochemical and structural characterization of SMYD2 activity toward p53 was recently used to propose $[LFM]_{-1}[K][AFYMSHRK]_{+1}[LYK]_{+2}$ as a consensus sequence for SMYD2 substrate specificity (30); however, whether the substrate specificity derived from such biochemical analyses encompasses all SMYD2 substrates in cells is unknown. In addition to substrates identified using biochemical approaches, an investigation into proteins that stably physically associate with SMYD2 led to the discovery of HSP90AA1-K615me1 as a Kme1 site in cells directly regulated by SMYD2 activity (31, 32). Despite the identification of potential substrates, the cellular and molecular biology of SMYD2 in oncogenesis remains largely understood—due, at least in part, to a lack of comprehensive characterization of endogenous SMYD2 activity in cells.

To address this challenge, we combined SILAC-based quantitative proteomics (33) with highly-specific and high-affinity pan-specific Kme1 antibody-based peptide enrichment followed by nano-LC-MS/MS analysis (8) to quantify

changes in Kme1 site abundance upon changes in SMYD2 activity. As a model system, we used the esophageal squamous cell carcinoma (ESCC) cell line KYSE-150, which harbors amplification of the *SMYD2* gene (27), and SMYD2 activity was modulated using either SMYD2 small molecule inhibitor LLY-507, shRNA-mediated knockdown, or stable overexpression of SMYD2. In total, we identified 1032 Kme1 sites in parental KYSE-150 cells and an additional set of 1861 Kme1 sites in KYSE-150 cells overexpressing SMYD2, comprising the most comprehensive and large-scale study of Kme1 sites reported to date. We reveal a set of 35 Kme1 sites that were down-regulated by knockdown or inhibition of SMYD2, of which only HSP90AA1-K615me1 is known. Furthermore, using biochemical methylation and mass spectrometry assays, we confirm the identification of four novel SMYD2 substrates—BTF3, PDAP1, as well as AHNK and AHNK2, which are two large proteins extensively methylated at multiple positions in their repetitive units by SMYD2. In addition, we report protein sequence motifs enriched in Kme1 sites upon SMYD2 overexpression that were also down-regulated upon perturbation of endogenous SMYD2, revealing unexpected sequence diversity in SMYD2 substrate recognition. Collectively, our data represents the first comprehensive and quantitative characterization of how protein methylation changes in response to methyltransferase perturbation and provides insights into SMYD2 substrates and specificity within the context of the cell.

EXPERIMENTAL PROCEDURES

SILAC Cell Culture Conditions—KYSE-150 cells were cultured in DMEM:F12 (1:1) media (Thermo Fisher Scientific, Grand Island, NY, PI88215) supplemented with 50 mg/L light (Sigma-Aldrich, St. Louis, MO L8662) or heavy ($^{13}\text{C}_6$, 99%; $^{15}\text{N}_2$, 99%; Cambridge Isotope Laboratories, Tewksbury, MA, CNLM-291-H-0.25g) lysine, 50 mg/L light (Sigma-Aldrich, A8094) or heavy ($^{13}\text{C}_6$, 99%; $^{15}\text{N}_2$, 99%; Cambridge Isotope Laboratories, CNLM-539-H-0.25g) arginine, 115 mg/L light proline (Sigma-Aldrich, P5607), and 10% dialyzed FBS (Thermo Scientific, PI-88440) at 37 °C and 5% CO_2 . KYSE-150 cells stably overexpressing SMYD2 (17) were cultured in the same medium and conditions, with the addition of 5 $\mu\text{g}/\text{ml}$ blasticidin (Thermo Scientific, R210-01).

Cell Treatments and Western Blotting—Cells were grown in light or heavy medium for six doublings and infected with lentivirus expressing a puromycin resistance gene as well as either control (in light medium) or SMYD2-targeted (in heavy medium) shRNA (Sigma-Aldrich) for 48 h, and then selected in the presence of 1.5 $\mu\text{g}/\text{ml}$ puromycin for an additional 48 h prior to harvest. KYSE-150 cells overexpressing SMYD2 (17) were cultured in light or heavy medium containing 5 $\mu\text{g}/\text{ml}$ blasticidin for 2 weeks, treated with 5 μM LLY-507 (in heavy medium) or DMSO (in light medium) for 24 h, and then harvested. Western blots were performed as previously described (17) using the following antibodies: anti-SMYD2 (generated in-house or Cell Signaling Technology, Danvers, MA, 9734 (D14H7)), anti- β -actin (Sigma-Aldrich, A2228), anti-p53 (Santa Cruz Biotechnology, Dallas, TX, SC-125 (DO1) or Cell Signaling Technology, 2524 (1C12)), and anti-p53-K370me1 (generated in-house).

Cell Lysis, Protein Digestion, and Strong Cation-Exchange (SCX) Fractionation—Cells were lysed in urea lysis buffer (8 M urea, 50 mM Tris-HCl, pH 8.3, 50 mM NaCl, and 1X Halt protease and phosphatase

inhibitor mixture (Thermo Scientific)) and sonicated. Protein lysates from the total light and heavy cell lysates were mixed at a 1:1 ratio, reduced using 10 mM DTT (Thermo Scientific), and alkylated using 50 mM iodoacetamide (Bio-Rad). Lysyl endopeptidase (Lys-C) (Wako Laboratory Chemicals, Richmond, VA) was added to the protein lysates at a ratio of 1:100 (w/w) for digestion at room temperature for 4 h. Urea was diluted to 1.5 M using 50 mM Tris-HCl (pH 8.3). Trypsin (Worthington Biochemical Corporation, Lakewood, NJ) was added at a ratio of 1:100 (w/w) for digestion overnight at room temperature. Peptides were desalted using Sep-Pak C18 cartridges (Waters, Milford, MA) as described previously (34), and lyophilized peptides were then dissolved in SCX Buffer A (10 mM KH_2PO_4 , 30% acetonitrile, pH 2.7), centrifuged at maximum speed for 5 min to remove insoluble material, and loaded onto a PolySULFOETHYL A™ column (9.4 mm I.D. \times 250 mm, PolyLC) using System Gold HPLC (Beckman Coulter, Fullerton, CA). Peptides were separated through a gradient of 100% Buffer A for 5 min, 0–25% Buffer B (10 mM KH_2PO_4 , 500 mM KCl, 30% acetonitrile, pH 2.7) for over 50 min, followed by 100% Buffer B for 5 min. Fractions were collected at 1 min intervals and combined into a total of eight fractions. Fractions were frozen in liquid nitrogen, lyophilized to remove acetonitrile, desalted using Sep-Pak C18 cartridges, and then lyophilized.

Cell Lysis and Subcellular Fractionation—Equal numbers of light and heavy cells were subject to subcellular fractionation as previously described (35, 36). Briefly, cells were lysed in ice-cold hypotonic lysis buffer (10 mM HEPES-KOH, pH 7.5, 10 mM KCl, 1.5 mM MgCl_2 , and $1\times$ Halt protease and phosphatase inhibitor mixture (Thermo Scientific)) and centrifuged at $225\times g$ for 5 min. Supernatant (cytoplasmic fraction) was transferred into a new tube, and the remaining nuclear pellet was resuspended in no-salt buffer (3 mM EDTA and 0.2 mM EGTA) and rotated for 30 min at 4 °C. Following centrifugation at $6500\times g$ for 3 min, supernatant and pellet were saved as soluble nuclear and chromatin fractions, respectively. Cytoplasmic and soluble nuclear fractions were mixed with a fourfold volume of chilled acetone overnight at -20°C , respectively, and protein pellets were dissolved in urea lysis buffer. Chromatin pellets were also dissolved in urea lysis buffer and sonicated.

Immunoaffinity Enrichment of Mono-methylated Peptides—Lyophilized peptides were dissolved in 1 ml PBS buffer and centrifuged at maximal speed for 5 min to remove any insoluble material. For SCX fractionation samples, supernatants of each fraction were incubated with 30 μg pan mono-methyl-lysine antibody (8) coupled to protein A Mag Sepharose beads (GE Healthcare, Piscataway, NJ) overnight at 4 °C. One-hundred micrograms of antibody was used for each subcellular fraction. Beads were washed three times with phosphate-buffered saline (PBS) buffer followed by three washes with Milli-Q water. Peptides were eluted from the beads twice using 100 μl 0.1% TFA, and fractions were then combined and lyophilized. Eluted peptides were desalted prior to LC-MS/MS analysis using in-house C18 STAGE tips as previously described (37).

Nano-LC-MS/MS Analysis and Protein Database Searches—Peptide samples were loaded onto a 75 μm I.D. \times 20 cm fused silica capillary column packed with Reprosil-Pur C18-AQ resin (3 μm ; Dr. Maisch GmbH, Germany) and resolved using an EASY-nLC 1000 HPLC system (Thermo Scientific) coupled in-line with a Q-Exactive (Thermo Scientific). The HPLC gradient consisted of 5–35% solvent B (A = 0.1% formic acid in water; B = 0.1% formic acid in acetonitrile) for 100–200 min, followed by 35% to 95% solvent B for 10 min, and then held at 95% solvent B for 10 min, with a constant flow-rate of 300 nl/min applied throughout. Full MS spectrum scans (m/z 350–1600) were performed at a resolution of 70,000 (at 200 m/z), and the 12 most intense ions were selected for MS/MS performed with high-energy collision dissociation (HCD) with normalized collision energy of 25 at a resolution of 17,500 (at 200 m/z). Automatic gain control

targets of full MS and MS/MS scans were 1×10^6 and 1×10^5 , respectively. Unassigned charge states and singly charged species were rejected, dynamic exclusion was set at 30 s, and lock mass calibration was implemented using polysiloxane ions 371.10123 and 445.12000. The mass spectrometry data, including annotated spectra for all mono-methylated peptides, was deposited to the ProteomeXchange Consortium (38) via the PRIDE partner repository with the dataset identifier PXD002405.

The pFind studio package (v3.10) was used for database searching and data analysis (39). Acquired MS/MS spectra were searched against a target-decoy version of the UniProt human protein database (May 16, 2014 release, 88,976 entries) consisting of forward and reverse protein sequences. A maximum of three trypsin miscleavage sites were allowed per peptide, and precursor ion and fragment ion tolerances were set to 5 ppm and 0.02 Da, respectively. Cysteine carbamidomethylation (+57.0215) was set as a static modification, with lysine mono-methylation (+14.0156), protein N-terminal acetylation (+42.0106), and methionine oxidation (+15.9949) set as dynamic modifications. The SILAC quantification module was enabled with lysine (+8.0142) and arginine (+10.0083). Peptide-spectrum matches with the same sequence, modifications, and isotope-labeling were treated as the same peptide. A peptide-level false discovery rate of 1.0% was applied to initially filter the search results (40), and independent false discovery rate calculations were subsequently performed for the mono-methyl-lysine peptides.

SILAC Ratios for Kme1 Sites, Data Normalization, and Statistical Analyses—SILAC quantification ratios for Kme1 sites were calculated as the median values of all Kme1 peptide ratios containing the same Kme1 site. For example, the SILAC ratio for the PDAP1-K126me1 site was calculated by averaging the individual ratios for the PDAP1 Kme1 peptides containing K126me1; specifically, the Y[Mox][Kme1]MHLAGK, Y[Mox][Kme1]MHLAGK, Y[Mox][Kme1]HLAGK, and Y[Mox][Kme1]HLAGK peptides (Mox = oxidized methionine). Kme1 site SILAC ratios were corrected for uneven total protein mixing of the heavy and light “input” protein lysates. The normalization factor to account for the uneven protein mixing was generated by running an aliquot of the combined peptide mixtures through LC-MS prior to anti-Kme1 antibody enrichment, and calculating the central value of the ratio distribution. Significance thresholds for differentially-regulated Kme1 peptides were defined as a fold-change ≥ 1.5 and p value ≤ 0.05 . The AKAP13-K1670me1 site was used as an internal reference point for calculating Kme1 site fold-changes and p values (Student's t test) for all Kme1 sites in the dataset.

SMYD2 Methylation Assays, SAH bioluminescence assay, and LC-MS/MS—Biochemical methylation assays using recombinant full-length human SMYD2 enzyme (purified in-house) were performed in 50 μl reactions overnight at 37 °C with shaking in 20 mM TRIZMA buffer (Sigma-Aldrich), pH 8.5 or 9.1, supplemented with 1 μM S-adenosyl-L-methionine (Sigma-Aldrich), 0.005% Surfact-Amps® 20 (Thermo Scientific), 4 mM Ultrapure DTT (Thermo Scientific), and 100 nM of the following His-tagged recombinant proteins: full-length BTF3 (Abcam, Cambridge, MA, 139205), full-length PDAP1 (Abcam, ab40188), full-length p53 (purified in-house), AHNAK (amino acids 4105–4634; Bio Basic Inc, Markham, ON), and AHNAK2 (amino acids 832–1491; Bio Basic Inc.). S-Adenosyl-L-homocysteine (SAH) levels were measured by bioluminescence using the MTase-GLO™ Reagent (Promega, Madison, WI) according to manufacturer instructions. Ten microliters of each methylation reaction was then denatured and reduced by incubation with an equal volume of 5 mM DTT in 100 mM ammonium bicarbonate containing 0.1% acid-labile detergent RapiGest (Waters) for 30 min at 60 °C, alkylated by incubation with iodoacetamide in 100 mM ammonium bicarbonate in the dark at room temperature for 30 min, and digested with the addition of 10 ng MS-grade trypsin (Thermo Scientific) in 10 μl of 100 mM ammonium

bicarbonate and overnight incubation at 37 °C. RapiGest was removed by the addition of 10 μ l 5% trifluoroacetic acid in 50% acetonitrile and incubation for 3 h at 60 °C. Three microliters of digested sample was injected into a nano-LC system (Dionex UltiMate 3000) coupled in-line with a Q-Exactive mass spectrometer (Thermo Scientific). Peptides were captured on a C18 trap column prior to washing, and eluted into a C18 resolving column (75 μ m I.D. \times 15 cm, Acclaim PepMap 100, Thermo Scientific) with a two-solvent gradient (A:B = 0.1% formic acid in water:0.1% formic acid in acetonitrile, 1% B per min). Mass spectrometric data were collected in full MS/AIF (all ion fragmentation) mode at 140,000/70,000 resolution for mass ranges of (200–2000)/(133–2000) Da. Primary MS data were initially searched using Pinpoint (Thermo Scientific) for all possible methylated peptides including multiple miscleavages and multiple methylations before the search was narrowed to focus on the list of lysine residues identified in the original SILAC experiments. Initial filtering criteria were as follows: (1) extracted ion chromatograms (XICs) of all four major isotopes within 10 ppm tolerance in primary MS spectra co-eluted and the relative abundance needed to match predicted abundance; (2) the observed charge distribution of the peptide needed to match the number of positive groups on the peptide; (3) the methylated peptide needed to migrate closely with the nonmethylated counterpart (if detected); and (4) the peptide sequence was supported by the AIF-MS2 fragmentation pattern of either the methylated or the nonmethylated peptide (or both depending on the abundance of each peptide). All peptides that passed these criteria were followed up with targeted MS2 for sequence confirmation and identification of the position of the mono-methylation within the peptide. The peak areas of the four major isotopes in the primary MS for the major charge states were used for quantitation of methylated and nonmethylated peptides.

RESULTS

Genetic and Pharmacological Disruption of SMYD2 Activity—To quantitatively measure changes in Kme1 sites following perturbation of SMYD2, we used SILAC (33) and the ESCC model cell line KYSE-150, which harbors elevated SMYD2 copy number and expression (27). We cultured cells in light- or heavy-labeled medium and generated lines that expressed control shRNA (“shControl”) or shRNA targeted against SMYD2 (“shSMYD2”), respectively. In addition, we generated cell lines that overexpressed SMYD2 (“SMYD2 overexpression”), cultured in either light- or heavy-labeled media, and then treated with DMSO or LLY-507, a selective and cell-active inhibitor of SMYD2 that we recently described (17), respectively. Consistent with SMYD2 loss of function, SMYD2 protein levels were depleted in heavy-labeled shSMYD2 cells relative to light-labeled shControl samples (Fig. 1A), and LLY-507 treatment strongly reduced the levels of p53 mono-methylated at lysine 370 (p53-K370me1) in SMYD2-overexpressing lines (Fig. 1B).

Identification of Kme1 Sites in ESCC Cells—Because SMYD2 is a protein lysine mono-methyltransferase (41), we used pan-specific mono-methyl-lysine polyclonal antibodies with excellent selectivity for Kme1 peptides that we previously validated as effective affinity reagents for peptide enrichment (8). Fig. 1C outlines a representative workflow for the replicate SCX-based HPLC peptide fractionation and Kme1 enrichment experiments. Briefly, protein extracted from a mixture

(1:1) of light-and heavy-labeled cells was digested into peptides, separated into eight fractions by SCX-HPLC, immunoprecipitated using mono-methyl-lysine antibodies, and identified and quantified using nano-LC-MS/MS analysis. Experiments were conducted as biological triplicate assays, involving the independent generation of both SMYD2 overexpression cell lines for LLY-507 treatment and SMYD2 shRNA knockdown cells. Information regarding SILAC peptide charge state and trypsin miscleavages, reproducibility of experimental replicates and Kme1 site identification, and other related information are found in [supplemental Figs. S1–S2](#). In total, we identified 1032 Kme1 sites in 740 distinct proteins in parental KYSE-150 cells and an additional 1861 Kme1 sites mapping to 1217 proteins in SMYD2-overexpressing cells ([supplemental Table S1](#)), comprising the largest and most comprehensive mono-methyl-lysine proteomic dataset reported to date (Fig. 1D).

To identify putative SMYD2 substrates, we searched for Kme1 sites down-regulated in response to SMYD2 knockdown or inhibition. We focused our quantitative analysis on a set of 273 robust and reproducible Kme1 sites (*i.e.* identified in at least two of three biological replicates) from 207 distinct proteins in parental KYSE-150 cells, of which the majority had not been reported (234 of the 273 Kme1 sites, or 85.7%). In SMYD2-overexpressing cells, there were 664 Kme1 sites in 498 distinct proteins that were reproducibly identified, of which we focused on a set of 198 Kme1 sites that overlapped with the parental cell line (Fig. 1E). We calculated SILAC-based quantification ratios (*i.e.* fold-changes) by averaging the Kme1 peptide spectra containing the same Kme1 site and normalized the ratios to the distributions of the total input digested sample (“normalized peptide ratio”). We defined differentially regulated Kme1 sites as those with normalized peptide ratios \geq 1.5-fold-change and used a significance threshold of *p* value \leq 0.05 relative to AKAP13-K1670me1, a Kme1 site presumptively regulated by the PKMT SETD7 (7) that was robustly identified with a near-1:1 SILAC-based quantification ratio across experimental replicates (Fig. 2A–2B, [supplemental Table S2](#), [supplemental Fig. S3](#)). Knockdown of SMYD2 in parental KYSE-150 cells resulted in the significant down-regulation of 66 Kme1 sites (66 of the 273 reproducible Kme1 sites, or 24.2%) and no sites were significantly up-regulated. Treatment of SMYD2-overexpressing cells with LLY-507 resulted in the down-regulation of 258 Kme1 sites (258 of the total 664 Kme1 sites, or 38.9%) with virtually all the remaining Kme1 sites showing no significant changes in abundance. HSP90AA1-K615me1, previously reported to be targeted by SMYD2 (31), was found in our analysis as a Kme1 site down-regulated by SMYD2 knockdown and inhibition, but there was no reliable supporting evidence for other Kme1 sites reported to be regulated by SMYD2, including RB1-K810me1 (42), PTEN-K313me1 (11), HSP90AB1-K531me1, or HSP90AB1-K574me1 (43). Moreover, Kme1 sites regulated by other PKMTs, such as

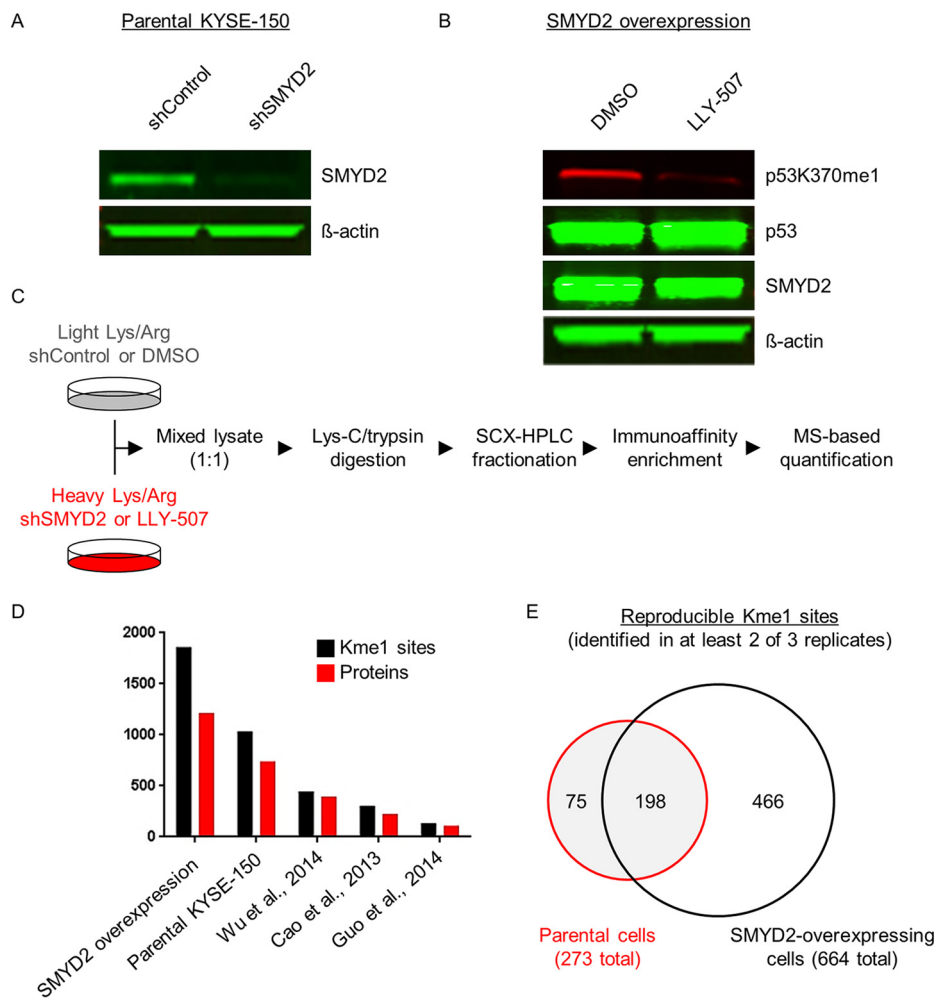


FIG. 1. Global identification of Kme1 sites using SILAC-based proteomics and Kme1 immunoaffinity enrichment. *A*, Western blot analysis of SMYD2 or β -actin in parental KYSE-150 cells infected with lentivirus expressing either control shRNA (shControl) or shRNA targeted against SMYD2 (shSMYD2). *B*, Western blot analysis of p53-K370me1, total p53, SMYD2, and β -actin following treatment of KYSE-150 cells overexpressing SMYD2 with either DMSO or LLY-507. *C*, Overview of the experimental approach used for SILAC-based peptide labeling and enrichment of lysine mono-methylated peptides in this study. *D*, Comparison of current and previously published datasets in terms of the number of Kme1 sites (black) and proteins with reported Kme1 sites (red). *E*, Venn diagram illustrating the number of reproducible Kme1 sites in parental cells and in SMYD2-overexpressing cells.

MAP3K2-K260me1 (44); WIZ-K967me1 and EEF1A1-K55me1 (25, 45), H3K27me1 (46), and H3K36me1 (47) (Fig. 2C) were unaffected by SMYD2 perturbation, suggesting that the quantitative changes in Kme1 site abundance that we observed were consistent with the potent and specific inhibition of SMYD2 activity.

In addition to HSP90AA1-K615me1, knockdown and inhibition of SMYD2 resulted in the down-regulation of 34 Kme1 sites (Fig. 2D–2E), including previously observed sites in CAPRIN1 (K675) (25) and PDAP1 (K126) (6, 7) that had not been ascribed to SMYD2 activity. Interestingly, multiple novel Kme1 sites in AHNK (K2589, K1205, and K1791) and AHNK2 (K939, K932, and K3414) were significantly down-regulated by SMYD2 knockdown and inhibition, indicating that multiple sites within these proteins are regulated SMYD2 activity. As with AKAP13-K1670me1, several other Kme1

sites, such as WIZ-K967me1, WIZ-K975me1, and OTUD4-K1040me1, showed little-to-no change in abundance in response to SMYD2 perturbation (Fig. 2E). Collectively, these data reveal a set of 35 SMYD2-regulated Kme1 sites, suggesting that SMYD2 targets a significantly wide spectrum of substrates in ESCC cells.

Candidate Kme1 Sites as Cellular Markers of Endogenous SMYD2 Activity—There are currently no robust or reliable cellular endpoints for the measurement of endogenous SMYD2 activity. Previously reported SMYD2-mediated Kme1 sites on p53 (10) and RB1 (13) require exogenous overexpression of either SMYD2 and/or the substrate protein in order to drive substrate methylation in cells. We reasoned that Kme1 sites that are commonly identified in proteomic datasets and in various cell lines are likely sufficiently abundant so as to serve as candidate markers of endogenous SMYD2 activity. Ac-

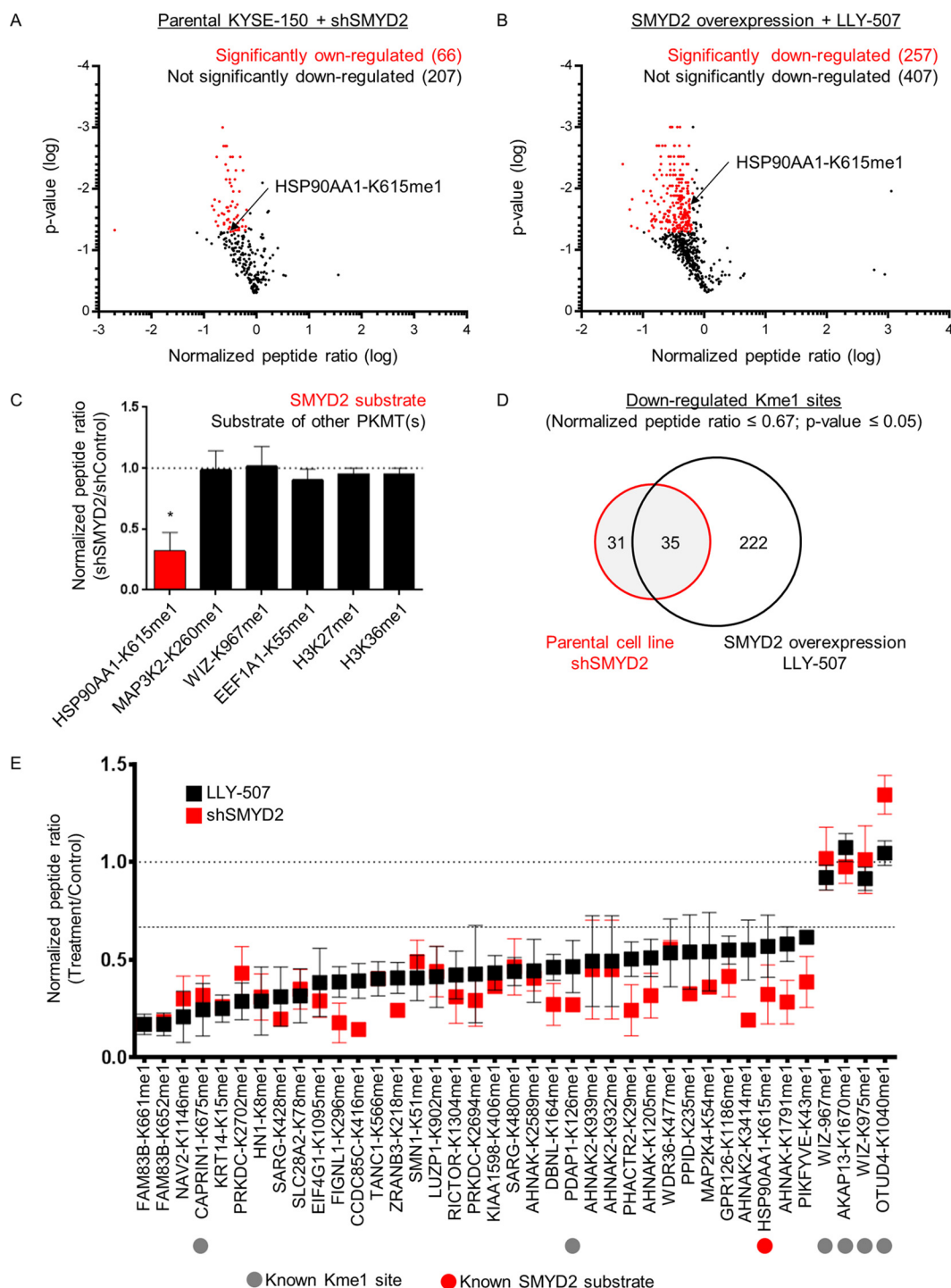


FIG. 2. Kme1 sites down-regulated following knockdown or inhibition of SMYD2. A, Volcano plot illustrating the normalized peptide ratio (shSMYD2/shControl) and p value distribution of Kme1 sites identified in control or SMYD2 knockdown cells. B, Volcano plot illustrating the (LLY-507/DMSO) and p value distribution of Kme1 sites identified in SMYD2-overexpressing KYSE-150 cells following treatment with LLY-507 or DMSO. C, Normalized peptide ratios of the SMYD2-dependent Kme1 site in HSP90AA1 (K615) upon SMYD2 knockdown as well as Kme1 sites known to be regulated by other PKMTs. D, Venn diagram illustrating the overlap in Kme1 sites (identified in at least two of three replicate experiments) identified in parental KYSE-150 and SMYD2-overexpressing cell lines. E, Venn diagram illustrating the number of Kme1 sites down-regulated by SMYD2 knockdown in parental KYSE-150 cells and by LLY-507 treatment in SMYD2-overexpressing cells. F, Normalized peptide ratios for Kme1 sites down-regulated in response to SMYD2 knockdown and to LLY-507 treatment.

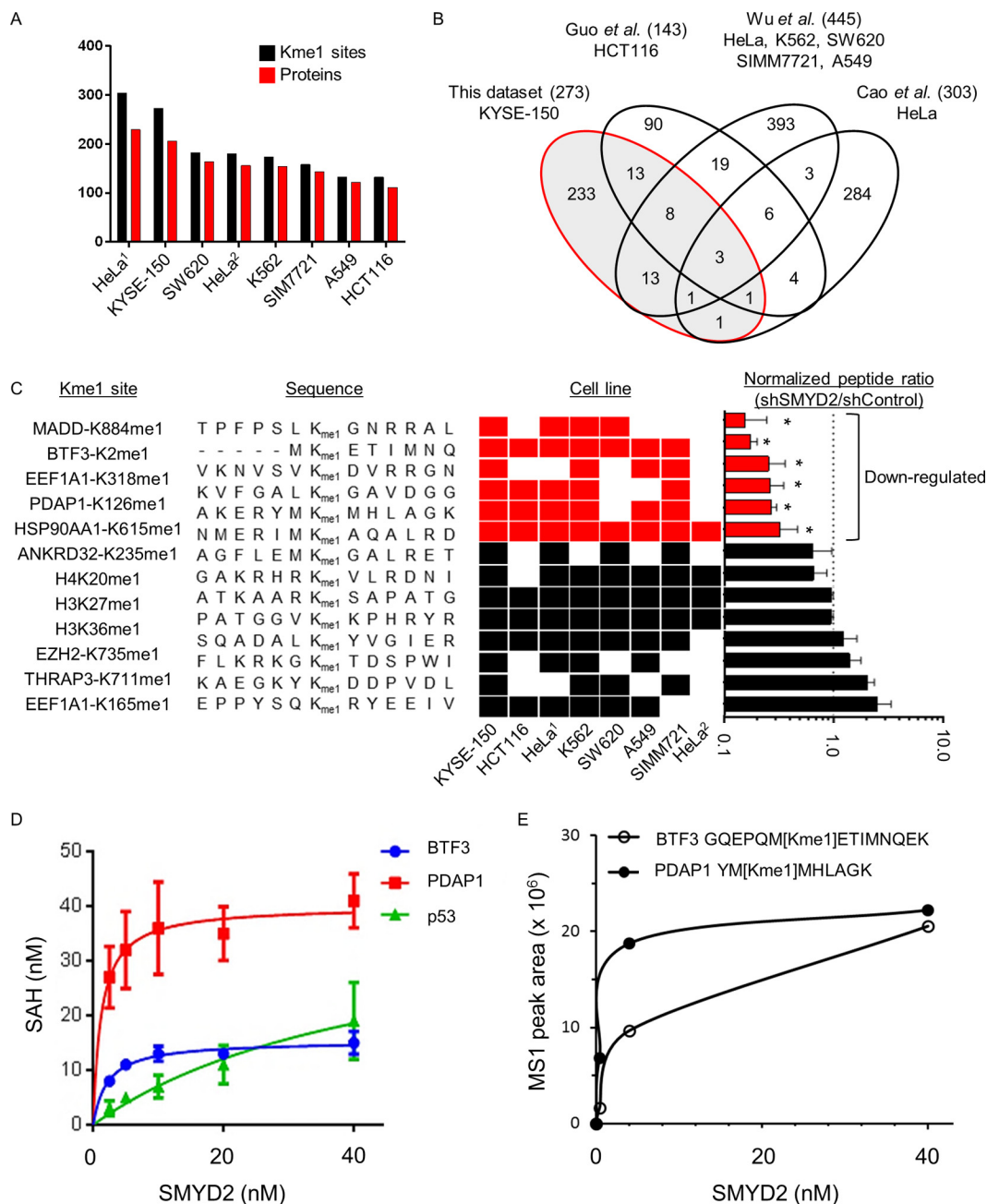


FIG. 3. SMYD2 knockdown down-regulates the abundance of Kme1 sites commonly identified in proteomic experiments and across a variety of cancer cell lines, including BTF3-K2 and PDAP1-K126. **A**, Number of Kme1 sites (red) and mono-methylated proteins (black) identified in the indicated cell lines. Only the 273 reproducible Kme1 sites in the parental KYSE-150 cell line are represented. **B**, The number of overlapping Kme1 sites identified in the indicated proteomic datasets among with the cell lines that were profiled in the respective studies. Specific Kme1 sites are listed in [Supplemental Table S3](#). **C**, Kme1 sites (and their sequences) that were identified in at least four of the seven cell lines profiles to date and their normalized SILAC peptide ratio in response to SMYD2 knockdown in parental KYSE-150 cells. * = *p* value ≤ 0.05 relative to AKAP13-K1670me1. **D**, Bioluminescence assays monitoring the production of SAH following incubation of 100 nM of protein substrate with the indicated concentrations of SMYD2. **E**, MS1 peak quantification of the indicated Kme1 peptide following biochemical methylation reactions using full-length recombinant BTF3 or PDAP1 with the indicated concentrations of SMYD2. 1, Cao *et al.*, 2013; 2, Wu *et al.*, 2014.

cordingly, we integrated the 273 reproducible Kme1 sites in KYSE-150 cells with other comparable proteomic Kme1 datasets (6–8), collectively encompassing seven cell lines and 1072 Kme1 sites (Fig. 3A). We then identified a set of 14 Kme1

sites that were consistently identified across the majority of cell lines and studies (*i.e.* identified in at least four of seven cell lines) (Fig. 3B–3C). We found that six of these 14 Kme1 sites, namely MADD-K884me1, BTF3-K2me1, EEF1A1-K318me1,

RPL5-K164me1, PDAP1-K126me1, and HSP90AA1-K615me1 were significantly down-regulated by SMYD2 knockdown in parental KYSE-150 cells (Fig. 3C).

Of these six proteins, we sought to further characterize SMYD2 activity toward BTF3 and PDAP1. To ensure accuracy of the SILAC quantification, we confirmed down-regulation of the SILAC ratios of the Kme1 peptides derived from these Kme1 sites by manual annotation of the MS1 and MS2 mass spectra (supplemental Figs. S4–S5). We then used Western blot analysis to ensure that SMYD2 knockdown did not result in significant protein-level down-regulation of either protein (supplemental Fig. S6), ruling out the possibility that changes in the SILAC ratios for these sites were the result of changes in total protein level. We then used biochemical methylation assays to monitor the activity of recombinant SMYD2 toward recombinant BTF3 or PDAP1 using bioluminescence assays to measure the conversion rate of the methyl-donor S-adenosyl-L-methionine (SAM) to S-adenosyl-L-homocysteine (SAH). Indeed, the use of BTF3 or PDAP1 as substrates, as with recombinant p53, resulted in increased SAH production in a manner proportional to SMYD2 concentration, confirming these two proteins as direct substrates of SMYD2 (Fig. 3D and supplemental Fig. S7A–S7B). Mass spectrometry analysis confirmed that recombinant BTF3 and PDAP1 were specifically mono-methylated at the K2 (or K46, depending on protein isoform) and K126 residues, respectively (supplemental Figs. S8–S9), consistent with the initial SILAC identification of these sites in cells (Fig. 3E). These data confirm BTF3 and PDAP1 as direct substrates of SMYD2 and highlight HSP90AA1-K615me1, BTF3-K2me1, and PDAP1-K126me1 as candidate markers for monitoring endogenous SMYD2 activity in cells.

Enriched Sequence Motifs in Kme1 Sites That Are Regulated by SMYD2 Activity—We next searched for sequence motifs that were enriched in the parental cell line-derived reproducible 273 Kme1 sites using the motif-X algorithm (48). We found that two sequence motifs, FK and LK, were significantly enriched in this peptide subset relative to the rest of the human proteome (Fig. 4A, supplemental Table S4). The FK and LK motifs were also enriched in the 664 Kme1 sites reproducibly identified in SMYD2-overexpressing cells, with the addition of LKGP, LKR, LKA, FKS, FKG, LKS, MK, and YK motifs. Of these 10 motifs, LKGP exhibited a particularly striking 122.4-fold enrichment relative to the human proteome (Fig. 4B), and, incidentally, all the Kme1 sites containing this motif mapped to either AHNK or AHNK2 (Fig. 2E). In addition, there were striking similarities, particularly at the –1 position, between the enriched motifs that we observed and the [LFM]₋₁[K][AFYMSHRK]₊₁[LYK]₊₂ substrate specificity derived from SMYD2 biochemical activity (30). However, we also observed a general lack of enrichment of specific amino acids at the +2 positions, suggesting that SMYD2 may recognize a more diverse collection of substrates in cells than originally predicted by biochemical and structural analyses.

Collectively, the enrichment of these sequence motifs was most likely the direct result of increased enzyme activity in SMYD2-overexpressing cells.

To assess whether the enriched sequence motifs identified in the SMYD2 overexpression dataset reflected the activity of endogenous SMYD2, we characterized the 273 Kme1 sites reproducibly identified in the parental KYSE-150 cell line according to the presence or absence of at least one of these motifs and compared the normalized peptide ratios of the two groups upon SMYD2 knockdown. We found that Kme1 sites containing an enriched sequence motif (167 of the total 273 Kme1 sites, or 61.2%) were significantly down-regulated in response to SMYD2 knockdown relative to the remaining Kme1 sites that did not contain an enriched sequence motif (106 of the total 273 Kme1 sites, or 38.8%) (Average normalized peptide ratio (“presence of motif”/“absence of motif”) between the two groups = 0.616, p value = $4E^{-9}$) (Fig. 4C). In addition, Kme1 sites containing LKS, LKGP, LKG, LKR, LKA, FKR, LK, or FK sequence motifs were also significantly down-regulated following knockdown of endogenous SMYD2 activity (Fig. 4D, supplemental Table S5). Although only a small number of Kme1 sites we identified in cells fit within the confines of the reported SMYD2 consensus specificity [LFM]₋₁[K][AFYMSHRK]₊₁[LYK]₊₂ (seven of the total 273 reproducible Kme1 sites, or 2.6%), these Kme1 sites were nonetheless significantly down-regulated as well, but not the Kme1 sites containing biochemical consensus sequences for the mono-methyltransferases SETD7 ([RK]₋₂[AST]₋₁[K]) or G9a/GLP ([R]₋₁[K]) (49, 50). Collectively, these data suggest the specific activity of endogenous SMYD2 toward Kme1 sites harboring these enriched motifs.

The Repetitive Units of AHNK and AHNK2 Are Uniquely Mono-methylated by SMYD2 at Multiple Sites—The majority of the mono-methylated proteins in our dataset contained either one or two Kme1 sites (197 of the total 206 mono-methylated proteins, or 95.6%). Because SMYD2 overexpression resulted in an increased number of Kme1 sites identified relative to the parental cell line (664 Kme1 sites and 273 Kme1 sites, respectively), we hypothesized that SMYD2 overexpression may have increased the number of Kme1 sites that occur in certain proteins. To test this possibility, we compared the number of Kme1 sites identified in each protein in parental cells relative to that of SMYD2-overexpressing cells. Indeed, the vast majority of proteins showed little-to-no change in the overall number of Kme1 sites identified per protein (i.e. one or two Kme1 sites per protein) with the exception of two proteins: AHNK and AHNK2. Overexpression of SMYD2 resulted in the identification of 10 additional Kme1 sites in AHNK2 (increased from seven sites to 17) and an additional 26 sites in AHNK (increased from 18 sites to 44) (Fig. 5A). We manually verified the accuracy of SILAC quantification for Kme1 peptides derived from these two proteins (supplemental Figs. S10–S11) and confirmed by immunofluorescence that, in the case of AHNK, SMYD2 knockdown

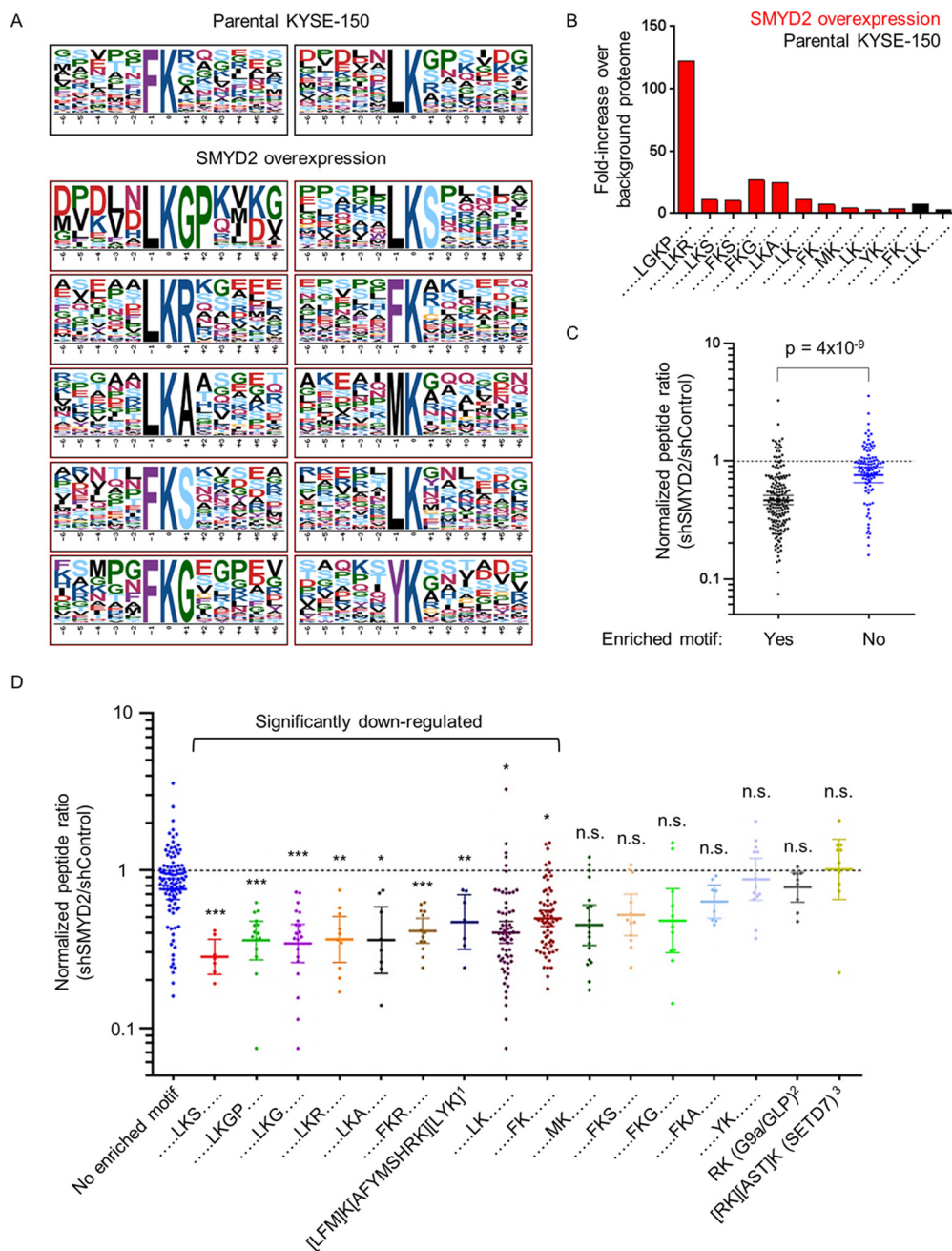


FIG. 4. Enrichment of known and novel specificity sequence motifs in SMYD2-targeted Kme1 sites. **A**, Sequence motifs enriched in the Kme1 sites identified in parental KYSE-150 cells or in SMYD2-overexpressing cells. **B**, Fold-increase in the occurrence of each enriched sequence motif compared with the background proteome based on motif-X analysis. See [Supplemental Table S4](#) for complete motif-X results. **C**, Normalized peptide ratios for Kme1 sites that contained an enriched sequence motif (black) and Kme1 sites that did not (blue). p value calculation is based on the normalized peptide ratios of Kme1 sites in each group using Student's t test. **D**, Normalized peptide ratios for Kme1 sites containing the indicated enriched sequence motif. p value calculations are based on comparison of the normalized peptide ratios of Kme1 sites in "No motif" using Student's t test. n.s., not significant; * p value ≤ 0.0005 ; ** p value ≤ 0.00005 ; *** p value ≤ 0.000005 . 1, Lanouette *et al.*, 2015; 2, Rathert *et al.*, 2008; 3, Dhayalan *et al.*, 2011.

did not affect total protein levels ([supplemental Fig. S12](#)). In both the parental and SMYD2-overexpressing cell lines, we found that the majority of Kme1 sites in AHNK and AHNK2 contained the LKGP sequence motif with, in some instances in AHNK2, leucine (L) at the -1 position replaced by isoleucine (I)

or valine (V) (Fig. 5B). Strikingly, approximately half the 271 occurrences of the LKGP sequence in the human proteome map to either AHNK or AHNK2 (135/271, 49.8%), suggesting that not only is the LKGP sequence motif enriched in Kme1 sites, but it is also highly over-represented in AHNK and AH-

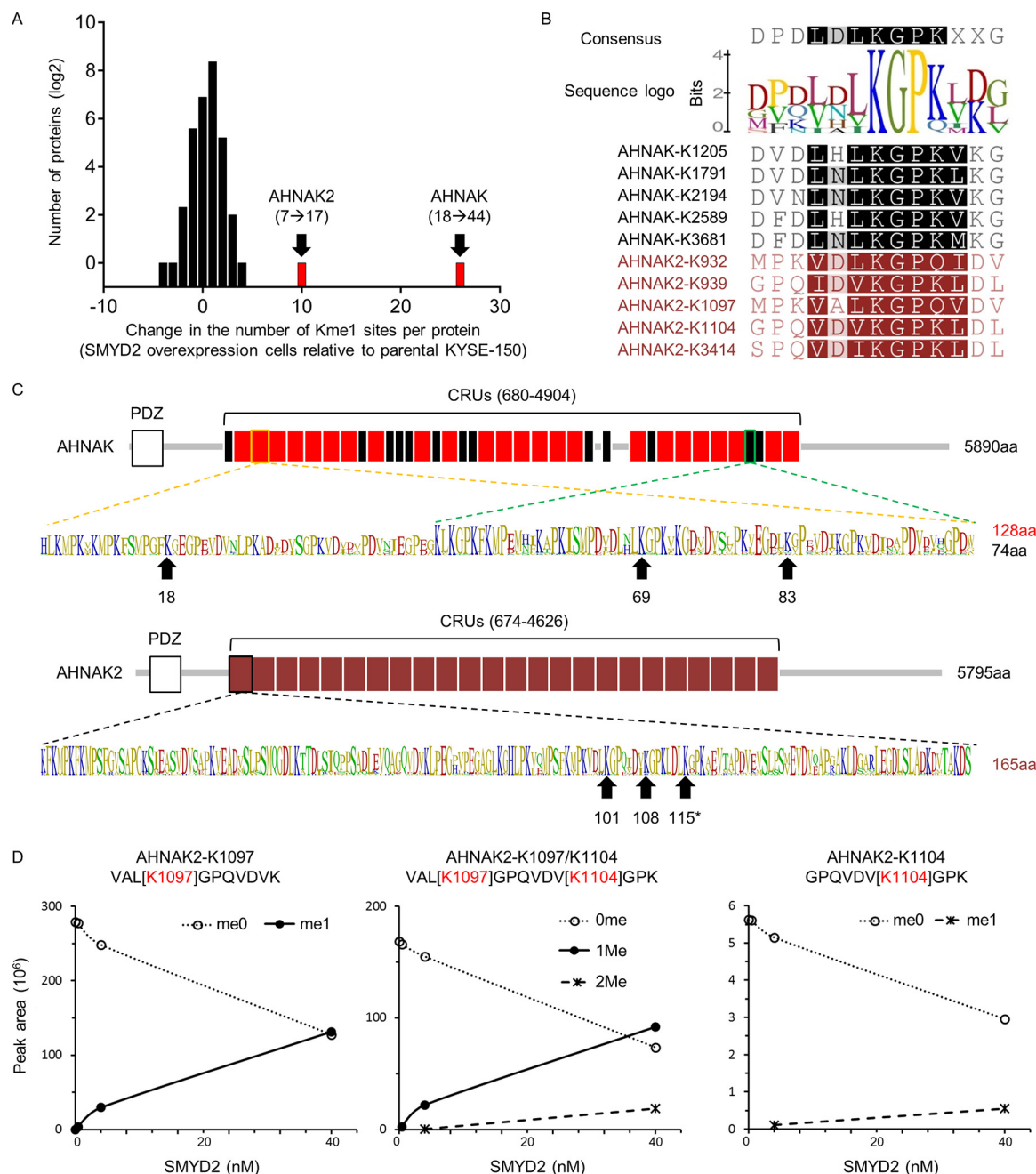


FIG. 5. AHNK and AHNK2 are mono-methylated at multiple sites by SMYD2. A, Changes in the number of Kme1 sites per protein identified in parental KYSE-150 cells and in SMYD2-overexpressing cells. Only reproducible Kme1 sites (i.e. identified in two of three experimental replicates) are included. AHNK and AHNK2 are highlighted in red. B, Protein sequence alignment of select Kme1 sites in AHNK and AHNK2 that were down-regulated by SMYD2 knockdown. C, Protein domain architecture of AHNK and AHNK2 and the positioning of the identified Kme1 sites within a representative consensus central repeat unit (CRU) of each protein. The asterisk indicates that this site ambiguously maps to AHNK as well. E, MS1 peak quantification of mono-methyl peptide products from biochemical methylation reactions of AHNK2-4CRU with the indicated concentration of recombinant SMYD2. Abbreviations: me0, unmodified peptide; me1, mono-methylated peptide; 1Me, single mono-methylation site; 2Me, double mono-methylation site.

NAK2 relative to the rest of the proteome. These findings show that AHNK and AHNK2 are unique among SMYD2 methylation substrates—and mono-methylated proteins in general—in that they are extensively mono-methylated at multiple sites and primarily at sites containing the LKGP motif.

AHNK and AHNK2 are described as giant proteins, given their estimated molecular weights of 629 and 617 kDa, respectively, and share similar protein architectures comprising three major protein domains: an N-terminal PDZ domain, a central region comprising several, sometimes degenerate, cen-

tral repeat units (CRUs), and a C-terminal domain. The central region of AHNAK comprises 24 128-mer CRUs that are interspersed with 13 shorter 76-mer CRUs, whereas AHNAK2 contains 24 contiguous 165-mer CRUs (Fig. 5C). Using the consensus 128-mer and 165-mer CRU sequences of AHNAK and AHNAK2, respectively, we mapped the Kme1 peptides from each protein to fixed positions within the consensus CRUs; specifically, lysines 18, 69, and 83 in the CRU of AHNAK, and lysines 101 and 108 in the CRU of AHNAK2 (Fig. 5C). Mass spectrometry data also suggested a potential additional Kme1 site exists at K115 in the AHNAK2 CRU; however, the peptides representing the AHNAK2-K115me1 site ambiguously mapped to AHNAK as well (supplemental Table S6). The repetitive nature of the CRUs of AHNAK and AHNAK2 and the multiple mapping of several Kme1 peptides within either AHNAK or AHNAK2 (see supplemental Table S6) prevent the precise determination of the number of SMYD2-mediated Kme1 sites in these two proteins based on mass spectrometry alone; however, after considering the mapping of peptides to multiple sites within AHNAK and AHNAK2, as well as the AHNAK2-K115me1 peptides ambiguously mapping to both proteins, we estimate that the number of Kme1 sites within each protein could range between a lower limit of 26 and 10 and an upper limit of 51 and 52 Kme1 sites in AHNAK and AHNAK2, respectively.

To confirm direct SMYD2-mediated mono-methylation of AHNAK and AHNAK2, we performed biochemical methylation assays using recombinant protein fragments consisting of four contiguous CRUs (4CRU) from either protein. As with BTF3 and PDAP1, incubation of recombinant SMYD2 with the four CRUs of either AHNAK or AHNAK2 resulted in increased levels of SAH in a manner proportional to the concentration of SMYD2 (supplemental Fig. S1). Using mass spectrometry, we confirmed that the SMYD2-dependent mono-methylation occurs at multiple LKGP sites at position 83 in the CRU of AHNAK, as well as multiple LKGP and VKGP sites at positions 101 and/or 108, respectively, in the CRU of AHNAK2 (Fig. 5D and supplemental Figs. S13–S14). Collectively, these data confirm that the CRUs of AHNAK and AHNAK2 are directly and extensively mono-methylated at multiple positions by SMYD2.

DISCUSSION

A Quantitative Proteomic Approach for the Characterization of PKMT Activity in Cells—Methods for broad spectrum identification of lysine methylation site signatures for PKMTs in cells is a critical bottleneck that limits the scope of investigations aimed at delineating their biological role(s). The quantitative proteomic approach that we described here, which coupled SILAC-based relative quantification, immunoaffinity enrichment, and nano-LC-MS/MS, resulted in the most comprehensive global proteomics study of mono-methyl-lysine reported to date and, with other Kme1 proteomic datasets, significantly expands the cellular landscape of mono-methyl-

lated lysine sites (6–8, 25). We also showed that quantitative proteomics using methyl-lysine immunoaffinity enrichment and genetic and pharmacological modulation of PKMT activity reveals unexpected and powerful global insights into PKMT substrates and substrate specificity that are not attainable by conventional approaches. Although SMYD2 was the PKMT selected for proof-of-concept, we emphasize that the application of this quantitative immunoaffinity proteomic approach to the study of other methyl-lysine regulatory proteins such as other PKMTs and lysine demethylases, is conceptually and technically straight-forward and can reveal impactful insights into enzyme activity.

Identification of Novel SMYD2 Substrates—The Kme1 proteomic dataset that we described here comprises the first comprehensive quantitative investigation into changes in Kme1 site abundance upon perturbation of the activity of a PKMT. Consistent with potent and selective disruption of SMYD2 activity, the known HSP90AA1-K615me1 site was down-regulated by SMYD2 knockdown, but not other Kme1 sites known to be regulated by other PKMTs (Fig. 2C). In addition, we reported 34 additional Kme1 sites down-regulated by knockdown or inhibition of SMYD2 (Fig. 2E). Furthermore, we showed that PDAP1-K126 and BTF3-K2 are directly mono-methylated by SMYD2 and, along with HSP90AA1-K615, show promise as potential markers for the direct monitoring of endogenous SMYD2 activity. Importantly, BTF3 and PDAP1 are broadly expressed across ESCC and other types of cancer cell lines, and the BTF3-K2me1 and PDAP1-K126me1 sites were also identified in cell lines that express relatively lower levels of SMYD2 (supplemental Fig. S16) (6), suggesting that over-expression of SMYD2 is not required for their detection in cells. As we recently reported (17), perturbation of SMYD2 had no significant quantitative effect on global levels of histone methylation, but rather affected the global levels of several non-histone Kme1 sites, as we demonstrated here, suggesting that SMYD2 primarily functions as a non-histone PKMT. In addition, there was no overlap between the Kme1 sites we identified in cells and the Kme1 sites (and tryptic Kme1 peptides) reported in biochemical methylation assays using recombinant proteins and artificial cell-based overexpression systems (11, 42, 43). Although we cannot confidently assert that these Kme1 sites do not occur in endogenous cellular contexts, we do emphasize the Kme1 sites we report here are of sufficient abundance for robust and reproducible monitoring and warrant attention in additional studies into the cellular activity of SMYD2.

Novel Insights into SMYD2 Substrate Recognition—We reported a collection of 10 enriched sequence motifs in Kme1 sites that were identified upon SMYD2 overexpression (Fig. 4A), and we showed that these Kme1 sites were also down-regulated in response to SMYD2 knockdown. These findings suggest that these motifs reflect the substrate specificity of SMYD2 in cells. The motif sequences we observed overlap with the recently reported [LFM]₋₁[K][AFYMSHRK]₊₁[LYK]₊₂

consensus specificity for SMYD2 (30), although only seven Kme1 sites that we identified in cells fit within the confines of this specificity signature. Moreover, this consensus specificity signature does not account for the novel Kme1 sites in BTF3, PDAP1, AHNAK, or AHNAK2 that we characterized in our study, suggesting that characterizing the activity of SMYD2 toward a single substrate (*i.e.* p53) may be insufficient to capture the breadth of its substrate specificity in cells; notably, the LKGP sequences in AHNAK and AHNAK2 likely represent a distinct mode of substrate recognition, one that merits further biochemical and structural characterization for potential new insights. Hence, we propose that our quantitative proteomics approach sheds insights into SMYD2 substrate specificity that can complement and inform systematic biochemical and structural analyses of enzyme activity.

New Frontiers to Explore for SMYD2 Biology—The identification of Kme1 sites that are regulated by SMYD2 activity in cells is intended to serve as a hypothesis-generating framework to enable further in-depth investigation into SMYD2 biology. The significance of protein mono-methylation and how these sites modulate the function(s) of these proteins needs to be characterized in detail. Although many of these putative SMYD2 substrates are largely functionally uncharacterized, there are nonetheless several substrates that raise intriguing possibilities for potential roles for SMYD2 in pro-growth signaling pathways. For example, there are now multiple lines of evidence supporting an oncogenic role for FAM83B, in which we identified SMYD2-mediated mono-methylation sites at K652 and K661 (51–53). We also identified a SMYD2-mediated mono-methylation site at K1304 in RICTOR, a component of the mTORC2 kinase complex (54) as well as in the intracellular domain of GPR126, an adhesion G-protein coupled receptor that bridges type-IV collagen in the extracellular matrix to intracellular cyclic AMP signaling (55). BTF3 is a component of the evolutionary-conserved nascent polypeptide-associated complex (NAC) that prevents the inappropriate recruitment of ribosomes to the endoplasmic reticulum (56, 57). The precise function of PDAP1 is unclear, but it has been reported to be highly up-regulated in the secretome of neoplastic gastric epithelial cells (58), a context in which SMYD2 activity was recently characterized (26). Finally, AHNAK and AHNAK2, which are uniquely and extensively directly mono-methylated by SMYD2, appear to possess diverse functionality (reviewed in (59)) with roles ranging from cell adhesion (60), cell signaling (61, 62), and tumor cell migration and invasion (63). Whether the extensive mono-methylation of the AHNAK and AHNAK2 CRUs by SMYD2, as well as the individual sites we identified in other SMYD2 substrates, regulates the function of these and other proteins remains to be determined.

Acknowledgments—We thank the pFind team at the Institute of Computing Technology for expert assistance in database searching.

* This work was supported in part by the National Institutes of

Health (GM110174 and AI118891) to B.A.G. This work was also supported by the Lilly Research Award Program (to B.A.G.). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

§ This article contains supplemental Figs. S1 to S15 and Tables S1 to S6.

|| To whom correspondence should be addressed: E-mail: nguyen_anh-quan_hannah@lilly.com and bgarci@mail.med.upenn.edu.

** These authors contributed equally to this work.

REFERENCES

- Bannister, A. J., and Kouzarides, T. (2011) Regulation of chromatin by histone modifications. *Cell Res.* **21**, 381–395
- Greer, E. L., and Shi, Y. (2012) Histone methylation: a dynamic mark in health, disease, and inheritance. *Nat. Rev. Genet.* **13**, 343–357
- Huang, J., and Berger, S. L. (2008) The emerging field of dynamic lysine methylation of non-histone proteins. *Curr. Opin. Genet. Dev.* **18**, 152–158
- Margueron, R., and Reinberg, D. (2010) Chromatin structure and the inheritance of epigenetic information. *Nat. Rev. Genet.* **11**, 285–296
- Su, I. H., and Tarakhovskiy, A. (2006) Lysine methylation and “signaling memory.” *Curr. Opin. Immunol.* **18**, 152–157
- Wu, Z., Cheng, Z., Sun, M., Wan, X., Liu, P., He, T., Tan, M., and Zhao, Y. (2015) A chemical proteomics approach for global analysis of lysine monomethylome profiling. *Mol. Cell. Proteomics* **14**, 329–339
- Guo, A., Gu, H., Zhou, J., Mulhern, D., Wang, Y., Lee, K. A., Yang, V., Aguiar, M., Kornhauser, J., Jia, X., Ren, J., Beausoleil, S. A., Silva, J. C., Vemulapalli, V., Bedford, M. T., and Comb, M. J. (2014) Immunoaffinity enrichment and mass spectrometry analysis of protein methylation. *Mol. Cell. Proteomics* **13**, 372–387
- Cao, X. J., Arnaudo, A. M., and Garcia, B. A. (2013) Large-scale global identification of protein lysine methylation *in vivo*. *Epigenetics* **8**, 477–485
- Hamamoto, R., Saloura, V., and Nakamura, Y. (2015) Critical roles of non-histone protein lysine methylation in human tumorigenesis. *Nat. Rev. Cancer* **15**, 110–124
- Huang, J., Perez-Burgos, L., Placek, B. J., Sengupta, R., Richter, M., Dorsey, J. A., Kubicek, S., Opravil, S., Jenuwein, T., and Berger, S. L. (2006) Repression of p53 activity by Smyd2-mediated methylation. *Nature* **444**, 629–632
- Nakakido, M., Deng, Z., Suzuki, T., Dohmae, N., Nakamura, Y., and Hamamoto, R. (2015) Dysregulation of AKT pathway by SMYD2-mediated lysine methylation on PTEN. *Neoplasia* **17**, 367–373
- Piao, L., Kang, D., Suzuki, T., Masuda, A., Dohmae, N., Nakamura, Y., and Hamamoto, R. (2014) The histone methyltransferase SMYD2 methylates PARP1 and promotes poly(ADP-ribosylation) activity in cancer cells. *Neoplasia* **16**, 257–264, 264 e252
- Saddic, L. A., West, L. E., Aslanian, A., Yates, J. R., 3rd, Rubin, S. M., Gozani, O., and Sage, J. (2010) Methylation of the retinoblastoma tumor suppressor by SMYD2. *J. Biol. Chem.* **285**, 37733–37740
- Arrowsmith, C. H., Bountra, C., Fish, P. V., Lee, K., and Schapira, M. (2012) Epigenetic protein families: a new frontier for drug discovery. *Nat. Rev. Drug Disc.* **11**, 384–400
- Barsyte-Lovejoy, D., Li, F., Oudhoff, M. J., Tatlock, J. H., Dong, A., Zeng, H., Wu, H., Freeman, S. A., Schapira, M., Senisterra, G. A., Kuznetsova, E., Marcellus, R., Allali-Hassani, A., Kennedy, S., Lambert, J. P., Couzens, A. L., Aman, A., Gingras, A. C., Al-Awar, R., Fish, P. V., Gerstenberger, B. S., Roberts, L., Benn, C. L., Grimley, R. L., Braam, M. J., Rossi, F. M., Sudol, M., Brown, P. J., Bunnage, M. E., Owen, D. R., Zaph, C., Vedadi, M., and Arrowsmith, C. H. (2014) (R)-PFI-2 is a potent and selective inhibitor of SETD7 methyltransferase activity in cells. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 12853–12858
- Konze, K. D., Ma, A., Li, F., Barsyte-Lovejoy, D., Parton, T., Macnevin, C. J., Liu, F., Gao, C., Huang, X. P., Kuznetsova, E., Rougie, M., Jiang, A., Pattenden, S. G., Norris, J. L., James, L. I., Roth, B. L., Brown, P. J., Frye, S. V., Arrowsmith, C. H., Hahn, K. M., Wang, G. G., Vedadi, M., and Jin, J. (2013) An orally bioavailable chemical probe of the Lysine Methyltransferases EZH2 and EZH1. *ACS Chem. Biol.* **8**, 1324–1334
- Nguyen, H., Allali-Hassani, A., Antonyamy, S., Chang, S., Chen, L. H., Curtis, C., Emtage, S., Fan, L., Gheyti, T., Li, F., Liu, S., Martin, J. R., Mendel, D., Olsen, J. B., Pelletier, L., Shatseva, T., Wu, S., Zhang, F. F.,

- Arrowsmith, C. H., Brown, P. J., Campbell, R. M., Garcia, B. A., Barsyte-Lovejoy, D., Mader, M., and Vedadi, M. (2015) LLY-507, a cell-active, potent, and selective inhibitor of protein lysine methyltransferase SMYD2. *J. Biol. Chem.*
18. Vedadi, M., Barsyte-Lovejoy, D., Liu, F., Rival-Gervier, S., Allali-Hassani, A., Labrie, V., Wigle, T. J., Dimaggio, P. A., Wasney, G. A., Siharheya, A., Dong, A., Tempel, W., Wang, S. C., Chen, X., Chau, I., Mangano, T. J., Huang, X. P., Simpson, C. D., Pattenden, S. G., Norris, J. L., Kireev, D. B., Tripathy, A., Edwards, A., Roth, B. L., Janzen, W. P., Garcia, B. A., Petronis, A., Ellis, J., Brown, P. J., Frye, S. V., Arrowsmith, C. H., and Jin, J. (2011) A chemical probe selectively inhibits G9a and GLP methyltransferase activity in cells. *Nat. Chem. Biol.* **7**, 566–574
19. Rikova, K., Guo, A., Zeng, Q., Possemato, A., Yu, J., Haack, H., Nardone, J., Lee, K., Reeves, C., Li, Y., Hu, Y., Tan, Z., Stokes, M., Sullivan, L., Mitchell, J., Wetzel, R., Macneill, J., Ren, J. M., Yuan, J., Bakalarski, C. E., Villen, J., Kornhauser, J. M., Smith, B., Li, D., Zhou, X., Gygi, S. P., Gu, T. L., Polakiewicz, R. D., Rush, J., and Comb, M. J. (2007) Global survey of phosphotyrosine signaling identifies oncogenic kinases in lung cancer. *Cell* **131**, 1190–1203
20. Rush, J., Moritz, A., Lee, K. A., Guo, A., Goss, V. L., Spek, E. J., Zhang, H., Zha, X. M., Polakiewicz, R. D., and Comb, M. J. (2005) Immunoaffinity profiling of tyrosine phosphorylation in cancer cells. *Nat. Biotechnol.* **23**, 94–101
21. Kim, W., Bennett, E. J., Huttlin, E. L., Guo, A., Li, J., Possemato, A., Sowa, M. E., Rad, R., Rush, J., Comb, M. J., Harper, J. W., and Gygi, S. P. (2011) Systematic and quantitative assessment of the ubiquitin-modified proteome. *Mol. Cell* **44**, 325–340
22. Choudhary, C., Kumar, C., Gnad, F., Nielsen, M. L., Rehman, M., Walther, T. C., Olsen, J. V., and Mann, M. (2009) Lysine acetylation targets protein complexes and coregulates major cellular functions. *Science* **325**, 834–840
23. Wagner, S. A., Beli, P., Weinert, B. T., Nielsen, M. L., Cox, J., Mann, M., and Choudhary, C. (2011) A proteome-wide, quantitative survey of *in vivo* ubiquitylation sites reveals widespread regulatory roles. *Mol. Cell. Proteomics* **10**, M111 013284
24. Liu, H., Galka, M., Mori, E., Liu, X., Lin, Y. F., Wei, R., Pittock, P., Voss, C., Dhami, G., Li, X., Miyaji, M., Lajoie, G., Chen, B., and Li, S. S. (2013) A method for systematic mapping of protein lysine methylation identifies functions for HP1beta in DNA damage response. *Mol. Cell* **50**, 723–735
25. Moore, K. E., Carlson, S. M., Camp, N. D., Cheung, P., James, R. G., Chua, K. F., Wolf-Yadlin, A., and Gozani, O. (2013) A general molecular affinity strategy for global detection and proteomic analysis of lysine methylation. *Mol. Cell* **50**, 444–456
26. Komatsu, S., Ichikawa, D., Hirajima, S., Nagata, H., Nishimura, Y., Kawaguchi, T., Miyamae, M., Okajima, W., Ohashi, T., Konishi, H., Shiozaki, A., Fujiwara, H., Okamoto, K., Tsuda, H., Imoto, I., Inazawa, J., and Otsuji, E. (2015) Overexpression of SMYD2 contributes to malignant outcome in gastric cancer. *Br. J. Cancer* **112**, 357–364
27. Komatsu, S., Imoto, I., Tsuda, H., Kozaki, K. I., Muramatsu, T., Shimada, Y., Aiko, S., Yoshizumi, Y., Ichikawa, D., Otsuji, E., and Inazawa, J. (2009) Overexpression of SMYD2 relates to tumor cell proliferation and malignant outcome of esophageal squamous cell carcinoma. *Carcinogenesis* **30**, 1139–1146
28. Sweis, R. F., Wang, Z., Algire, M., Arrowsmith, C. H., Brown, P. J., Chiang, G. G., Guo, J., Jakob, C. G., Kennedy, S., Li, F., Maag, D., Shaw, B., Soni, N. B., Vedadi, M., and Pappano, W. N. (2015) Discovery of A-893, a new cell-active benzoxazinone inhibitor of lysine methyltransferase SMYD2. *ACS Med. Chem. Lett.* **6**, 695–700
29. Ferguson, A. D., Larsen, N. A., Howard, T., Pollard, H., Green, I., Grande, C., Cheung, T., Garcia-Arenas, R., Cowen, S., Wu, J., Godin, R., Chen, H., and Keen, N. (2011) Structural basis of substrate methylation and inhibition of SMYD2. *Structure* **19**, 1262–1273
30. Lanouette, S., Davey, J. A., Elisma, F., Ning, Z., Figeys, D., Chica, R. A., and Couture, J. F. (2015) Discovery of substrates for a SET domain lysine methyltransferase predicted by multistate computational protein design. *Structure* **23**, 206–215
31. Abu-Farha, M., Lanouette, S., Elisma, F., Tremblay, V., Butson, J., Figeys, D., and Couture, J. F. (2011) Proteomic analyses of the SMYD family interactomes identify HSP90 as a novel target for SMYD2. *J. Mol. Cell Biol.* **3**, 301–308
32. Donlin, L. T., Andresen, C., Just, S., Rudensky, E., Pappas, C. T., Kruger, M., Jacobs, E. Y., Unger, A., Ziesenis, A., Dobenecker, M. W., Voelkel, T., Chait, B. T., Gregorio, C. C., Rottbauer, W., Tarakhovsky, A., and Linke, W. A. (2012) Smyd2 controls cytoplasmic lysine methylation of Hsp90 and myofilament organization. *Genes Dev.* **26**, 114–119
33. Ong, S. E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A., and Mann, M. (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics* **1**, 376–386
34. Villen, J., and Gygi, S. P. (2008) The SCX/IMAC enrichment approach for global phosphorylation analysis by mass spectrometry. *Nat. Protoc.* **3**, 1630–1638
35. Shechter, D., Dormann, H. L., Allis, C. D., and Hake, S. B. (2007) Extraction, purification, and analysis of histones. *Nat. Protoc.* **2**, 1445–1457
36. Andersen, J. S., Lyon, C. E., Fox, A. H., Leung, A. K., Lam, Y. W., Steen, H., Mann, M., and Lamond, A. I. (2002) Directed proteomic analysis of the human nucleolus. *Curr. Biol.* **12**, 1–11
37. Rappsilber, J., Ishihama, Y., and Mann, M. (2003) Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal. Chem.* **75**, 663–670
38. Vizcaino, J. A., Deutsch, E. W., Wang, R., Csordas, A., Reisinger, F., Rios, D., Dienes, J. A., Sun, Z., Farrah, T., Bandeira, N., Binz, P. A., Xenarios, I., Eisenacher, M., Mayer, G., Gatto, L., Campos, A., Chalkley, R. J., Kraus, H. J., Albar, J. P., Martinez-Bartolome, S., Apweiler, R., Omenn, G. S., Martens, L., Jones, A. R., and Hermjakob, H. (2014) ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nat. Biotechnol.* **32**, 223–226
39. Li, D., Fu, Y., Sun, R., Ling, C. X., Wei, Y., Zhou, H., Zeng, R., Yang, Q., He, S., and Gao, W. (2005) pFind: a novel database-searching software system for automated peptide and protein identification via tandem mass spectrometry. *Bioinformatics* **21**, 3049–3050
40. Elias, J. E., Haas, W., Faherty, B. K., and Gygi, S. P. (2005) Comparative evaluation of mass spectrometry platforms used in large-scale proteomics investigations. *Nat. Methods* **2**, 667–675
41. Wang, L., Li, L., Zhang, H., Luo, X., Dai, J., Zhou, S., Gu, J., Zhu, J., Atadja, P., Lu, C., Li, E., and Zhao, K. (2011) Structure of human SMYD2 protein reveals the basis of p53 tumor suppressor methylation. *J. Biol. Chem.* **286**, 38725–38737
42. Cho, H. S., Hayami, S., Toyokawa, G., Maejima, K., Yamane, Y., Suzuki, T., Dohmae, N., Kogure, M., Kang, D., Neal, D. E., Ponder, B. A., Yamaue, H., Nakamura, Y., and Hamamoto, R. (2012) RB1 methylation by SMYD2 enhances cell cycle progression through an increase of RB1 phosphorylation. *Neoplasia* **14**, 476–486
43. Hamamoto, R., Toyokawa, G., Nakakido, M., Ueda, K., and Nakamura, Y. (2014) SMYD2-dependent HSP90 methylation promotes cancer cell proliferation by regulating the chaperone complex formation. *Cancer Lett.* **351**, 126–133
44. Mazur, P. K., Reynoird, N., Khatri, P., Jansen, P. W., Wilkinson, A. W., Liu, S., Barbash, O., Van Aller, G. S., Huddleston, M., Dhanak, D., Tummino, P. J., Kruger, R. G., Garcia, B. A., Butte, A. J., Vermeulen, M., Sage, J., and Gozani, O. (2014) SMYD3 links lysine methylation of MAP3K2 to Ras-driven cancer. *Nature* **510**, 283–287
45. Rathert, P., Dhayalan, A., Murakami, M., Zhang, X., Tamas, R., Jurkowska, R., Komatsu, Y., Shinkai, Y., Cheng, X., and Jeltsch, A. (2008) Protein lysine methyltransferase G9a acts on non-histone targets. *Nat. Chem. Biol.* **4**, 344–346
46. Margueron, R., and Reinberg, D. (2011) The Polycomb complex PRC2 and its mark in life. *Nature* **469**, 343–349
47. Wagner, E. J., and Carpenter, P. B. (2012) Understanding the language of Lys36 methylation at histone H3. *Nat. Rev. Mol. Cell Biol.* **13**, 115–126
48. Schwartz, D., Chou, M. F., and Church, G. M. (2009) Predicting protein post-translational modifications using meta-analysis of proteome scale data sets. *Mol. Cell. Proteomics* **8**, 365–379
49. Dhayalan, A., Kudithipudi, S., Rathert, P., and Jeltsch, A. (2011) Specificity analysis-based identification of new methylation targets of the SET7/9 protein lysine methyltransferase. *Chem. Biol.* **18**, 111–120
50. Couture, J. F., Collazo, E., Hauk, G., and Trievel, R. C. (2006) Structural basis for the methylation site specificity of SET7/9. *Nat. Struct. Mol. Biol.* **13**, 140–146
51. Cipriano, R., Bryson, B. L., Miskimen, K. L., Bartel, C. A., Hernandez-Sanchez, W., Bruntz, R. C., Scott, S. A., Lindsley, C. W., Brown, H. A.,

- and Jackson, M. W. (2014) Hyperactivation of EGFR and downstream effector phospholipase D1 by oncogenic FAM83B. *Oncogene* **33**, 3298–3306
52. Cipriano, R., Graham, J., Miskimen, K. L., Bryson, B. L., Bruntz, R. C., Scott, S. A., Brown, H. A., Stark, G. R., and Jackson, M. W. (2012) FAM83B mediates EGFR- and RAS-driven oncogenic transformation. *J. Clin. Invest.* **122**, 3197–3210
 53. Cipriano, R., Miskimen, K. L., Bryson, B. L., Foy, C. R., Bartel, C. A., and Jackson, M. W. (2013) FAM83B-mediated activation of PI3K/AKT and MAPK signaling cooperates to promote epithelial cell transformation and resistance to targeted therapies. *Oncotarget* **4**, 729–738
 54. Oh, W. J., and Jacinto, E. (2011) mTOR complex 2 signaling and functions. *Cell Cycle* **10**, 2305–2316
 55. Paavola, K. J., Sidik, H., Zuchero, J. B., Eckart, M., and Talbot, W. S. (2014) Type IV collagen is an activating ligand for the adhesion G protein-coupled receptor GPR126. *Sci. Signal.* **7**, ra76
 56. Zhang, Y., Berndt, U., Golz, H., Tais, A., Oellerer, S., Wolfle, T., Fitzke, E., and Rospert, S. (2012) NAC functions as a modulator of SRP during the early steps of protein targeting to the endoplasmic reticulum. *Mol. Biol. Cell* **23**, 3027–3040
 57. Gamerding, M., Hanebuth, M. A., Frickey, T., and Deuerling, E. (2015) The principle of antagonism ensures protein targeting specificity at the endoplasmic reticulum. *Science* **348**, 201–207
 58. Marimuthu, A., Subbannayya, Y., Sahasrabudhe, N. A., Balakrishnan, L., Syed, N., Sekhar, N. R., Katte, T. V., Pinto, S. M., Srikanth, S. M., Kumar, P., Pawar, H., Kashyap, M. K., Maharudraiah, J., Ashktorab, H., Smoot, D. T., Ramaswamy, G., Kumar, R. V., Cheng, Y., Meltzer, S. J., Roa, J. C., Chaerkady, R., Prasad, T. S., Harsha, H. C., Chatterjee, A., and Pandey, A. (2013) SILAC-based quantitative proteomic analysis of gastric cancer secretome. *Proteomics Clin. Appl.* **7**, 355–366
 59. Davis, T. A., Loos, B., and Engelbrecht, A. M. (2014) AHNAK: the giant jack of all trades. *Cell. Signal.* **26**, 2683–2693
 60. De Seranno, S., Benaud, C., Assard, N., Khediri, S., Gerke, V., Baudier, J., and Delphin, C. (2006) Identification of an AHNAK binding motif specific for the Annexin2/S100A10 tetramer. *J. Biol. Chem.* **281**, 35030–35038
 61. Lee, I. H., Sohn, M., Lim, H. J., Yoon, S., Oh, H., Shin, S., Shin, J. H., Oh, S. H., Kim, J., Lee, D. K., Noh, D. Y., Bae, D. S., Seong, J. K., and Bae, Y. S. (2014) Ahnak functions as a tumor suppressor via modulation of TGFbeta/Smad signaling pathway. *Oncogene* **33**, 4675–4684
 62. Lee, I. H., Lim, H. J., Yoon, S., Seong, J. K., Bae, D. S., Rhee, S. G., and Bae, Y. S. (2008) Ahnak protein activates protein kinase C (PKC) through dissociation of the PKC-protein phosphatase 2A complex. *J. Biol. Chem.* **283**, 6312–6320
 63. Sudo, H., Tsuji, A. B., Sugyo, A., Abe, M., Hino, O., and Saga, T. (2014) AHNAK is highly expressed and plays a key role in cell migration and invasion in mesothelioma. *Int. J. Oncol.* **44**, 530–538